**Описание проекта**

Допустим, вы работаете в добывающей компании «ГлавРосГосНефть». Нужно решить, где бурить новую скважину.

Шаги для выбора локации обычно такие:

* В избранном регионе собирают характеристики для скважин: качество нефти и объём её запасов;
* Строят модель для предсказания объёма запасов в новых скважинах;
* Выбирают скважины с самыми высокими оценками значений;
* Определяют регион с максимальной суммарной прибылью отобранных скважин.

Вам предоставлены пробы нефти в трёх регионах. Характеристики для каждой скважины в регионе уже известны. Постройте модель для определения региона, где добыча принесёт наибольшую прибыль. Проанализируйте возможную прибыль и риски техникой *Bootstrap.*

Данные геологоразведки трёх регионов находятся в файлах:

* /datasets/geo\_data\_0
* /datasets/geo\_data\_1
* /datasets/geo\_data\_2
* *id* — уникальный идентификатор скважины;
* *f0, f1, f2* — три признака точек (неважно, что они означают, но сами признаки значимы);

*product* — объём запасов в скважине (тыс. баррелей).

**Наш план:**

1. Загрузим и подготовим данные. Поясним порядок действий.

2. Обучим и проверим модель для каждого региона:

3. Подготовим данные к расчёту прибыли.

4. Напишем функцию для расчёта прибыли по выбранным скважинам и предсказаниям модели.

5. Посчитаем риски и прибыль для каждого региона.

6. Напишем выводы: предложим регион для разработки скважин и обоснуем выбор.

* Для обучения модели подходит только линейная регрессия (остальные — недостаточно предсказуемые).

Данные синтетические: детали контрактов и характеристики месторождений не разглашаются.

В этом проекте будут использованы следующие библиотеки:

pandas, seaborn, scikit-learn (sklearn).